Vol. 83, No. 17

Mixed Infection and the Genesis of Influenza Virus Diversity[▽]

Elodie Ghedin,[1]* Adam Fitch,[1] Alex Boyne,[2] Sara Griesemer,[3] Jay DePasse,[1] Jayati Bera,[2]
Xu Zhang,[1] Rebecca A. Halpin,[2] Marita Smit,[4] Lance Jennings,[4] Kirsten St. George,[3]
Edward C. Holmes,[5,6] and David J. Spiro[2]

*University of Pittsburgh School of Medicine, Departments of Medicine and Computational Biology, Pittsburgh, Pennsylvania 15261[1];
J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850[2]; Wadsworth Center, New York State Department of Health,
Albany, New York 12201[3]; Canterbury Health Laboratories, Christchurch, New Zealand[4]; The Pennsylvania State University,
University Park, Pennsylvania 16802[5]; and Fogarty International Center, National Institutes of Health,
Bethesda, Maryland 20892[6]*

The emergence of viral infections with potentially devastating consequences for human health is highly dependent on their underlying evolutionary dynamics. One likely scenario for an avian influenza virus, such as A/H5N1, to evolve to one capable of human-to-human transmission is through the acquisition of genetic material from the A/H1N1 or A/H3N2 subtypes already circulating in human populations. This would require that viruses of both subtypes coinfect the same cells, generating a mixed infection, and then reassort. Determining the nature and frequency of mixed infection with influenza virus is therefore central to understanding the emergence of pandemic, antigenic, and drug-resistant strains. To better understand the potential for such events, we explored patterns of intrahost genetic diversity in recently circulating strains of human influenza virus. By analyzing multiple viral genome sequences sampled from individual influenza patients we reveal a high level of mixed infection, including diverse lineages of the same influenza virus subtype, drug-resistant and -sensitive strains, those that are likely to differ in antigenicity, and even viruses of different influenza virus types (A and B). These results reveal that individuals can harbor influenza viruses that differ in major phenotypic properties, including those that are antigenically distinct and those that differ in their sensitivity to antiviral agents.

Influenza viruses (family *Orthomyxoviridae*) possess a negative-strand segmented RNA genome and enveloped virions. Genetic diversity in influenza virus is the result of a high rate of mutation associated with replication using low-fidelity RNA polymerase and of the reshuffling (or reassortment) of segments among coinfecting strains. Although the 13.5-kb genome of influenza A virus is composed of eight segments coding for 11 known proteins, these viruses are typically categorized by their two surface antigens, hemagglutinin (HA), of which there are 16 subtypes (H1 to H16), and neuraminidase (NA), of which there are 9 (N1 to N9) (9). All known subtypes are present in aquatic birds of the orders *Anseriformes* and *Charadriiformes*, and a smaller number circulate in some mammalian species. The HA plays a major role in the attachment of the virus to the host cell surface by binding to the sialic acid moiety of host receptors and facilitating the fusion of the viral envelope with host cell membranes. It is also the major viral antigen against which neutralizing antibodies are directed. The NA is important for mobility of the virions by cleaving the sialic acid residues from the viral hemagglutinin, which facilitates both entry of the virus into the cell and release of the viruses during budding (11).

Most discussions of influenza virus evolution have focused on the process of antigenic drift in which mutations accumulate—most likely by natural selection—in the antigenic sites of the HA and NA, thereby allowing evasion of the host populations' acquired immunity to previously circulating strains. Such antigenic variation occurs primarily in the HA1 domain and is clustered into five main epitope regions (19, 20, 22). Although antigenic drift clearly plays a key role in the seasonal evolution of influenza A virus, recent studies making use of large data sets generated by the Influenza Genome Sequencing Project (IGSP) suggest that reassortment may also be important in the generation of antigenically novel isolates by placing diverse HAs in compatible genetic backgrounds (6, 8, 10, 14).

Segment reassortment is also central to the process of cross-species transmission and emergence of pandemic influenza virus. In particular, the segmented nature of the influenza virus genome allows reassortment of gene segments to occur between diverse influenza A virus strains when they coinfect a single host, including those derived from different species. This can result in subtle changes within a subtype, or dramatic changes that occur when different subtypes mix, leading to the generation of novel viruses expressing surface glycoproteins to which a specific host immune system has little if any serological cross-reactivity. Such antigenic shift is believed to have led to the emergence of global human influenza A virus pandemics in 1957 (A/H2N2) and in 1968 (A/H3N2), with new segments ultimately derived from the avian reservoir pool reassorting into human influenza viruses (17).

Given the potential for emerging viruses such as influenza virus to adversely affect the health of human and other animal populations, it is essential to determine the factors that allow viruses to acquire the mutations they need to adapt to new host

---

* Corresponding author. Mailing address: Division of Infectious Diseases, University of Pittsburgh School of Medicine, 3550 Terrace Street, Room 830, Pittsburgh, PA 15261. Phone: (412) 383-5850. Fax: (412) 383-5851. E-mail: elg21@pitt.edu.

TABLE 1. Influenza A virus segment clones from patients NZ094 and WW537 and GenBank accession numbers

Organism	Segment	Blinded no.	Clone name	GenBank accession no.
Influenza A virus Canterbury/200/2004 (H3N2)	1 (PB2)	NIGSP-NZ-00094.1a	NZ094P32S1C004	CY039886
		NIGSP-NZ-00094.1b	NZ094P32S1C008	CY039887
		NIGSP-NZ-00094.1c	NZ094P32S1C002	CY039888
		NIGSP-NZ-00094.1d	NZ094P32S1C003	CY039889
		NIGSP-NZ-00094.1e	NZ094P32S1C005	CY039890
		NIGSP-NZ-00094.1f	NZ094P32S1C006	CY039891
	2 (PB1)	NIGSP-NZ-00094.2a	NZ094P32S2C026	CY039892
		NIGSP-NZ-00094.2b	NZ094P11S2C006	CY039939
		NIGSP-NZ-00094.2c	NZ094P11S2C018	CY039940
		NIGSP-NZ-00094.2d	NZ094P32S2C021	CY039941
		NIGSP-NZ-00094.2e	NZ094P32S2C024	CY039942
		NIGSP-NZ-00094.2f	NZ094P32S2C028	CY039943
	3 (PA)	NIGSP-NZ-00094.3a	NZ094P32S3C101	CY039944
		NIGSP-NZ-00094.3b	NZ094P32S3C102	CY039945
		NIGSP-NZ-00094.3c	NZ094P32S3C104	CY039946
		NIGSP-NZ-00094.3d	NZ094P32S3C201	CY039947
		NIGSP-NZ-00094.3e	NZ094P32S3C205	CY039948
		NIGSP-NZ-00094.3f	NZ094P32S3C302	CY039949
		NIGSP-NZ-00094.3g	NZ094P32S3C303	CY039950
		NIGSP-NZ-00094.3h	NZ094P32S3C401	CY039951
		NIGSP-NZ-00094.3i	NZ094P32S3C405	CY039952
		NIGSP-NZ-00094.3j	NZ094P32S3C406	CY039953
		NIGSP-NZ-00094.3k	NZ094P32S3C503	CY039954
		NIGSP-NZ-00094.3l	NZ094P32S3C504	CY039955
		NIGSP-NZ-00094.3m	NZ094P32S3C602	CY039956
		NIGSP-NZ-00094.3n	NZ094P32S3C603	CY039957
		NIGSP-NZ-00094.3o	NZ094P32S3C604	CY039958
		NIGSP-NZ-00094.3p	NZ094P32S3C605	CY039959
	4 (HA)	NIGSP-NZ-00094.4a	NZ094P32S4Cx06	CY039960
		NIGSP-NZ-00094.4b	NZ094P32S4Cx04	CY039961
		NIGSP-NZ-00094.4c	NZ094P32S4Cx03	CY039962
		NIGSP-NZ-00094.4d	NZ094P32S4Cx02	CY039963
		NIGSP-NZ-00094.4e	NZ094P32S4Cx01	CY039964
		NIGSP-NZ-00094.4f	NZ094P32S4C007	CY039965
		NIGSP-NZ-00094.4g	NZ094P32S4C006	CY039966
	5 (NP)	NIGSP-NZ-00094.5a	NZ094P11S5C002	CY039967
		NIGSP-NZ-00094.5b	NZ094P11S5C005	CY039968
		NIGSP-NZ-00094.5c	NZ094P32S5C001	CY039969
		NIGSP-NZ-00094.5d	NZ094P32S5C004	CY039970
	6 (NA)	NIGSP-NZ-00094.6a	NZ094P32S6C001	CY039971
		NIGSP-NZ-00094.6b	NZ094P32S6C003	CY039972
		NIGSP-NZ-00094.6c	NZ094P32S6C005	CY039973
	7 (M)	NIGSP-NZ-00094.7a	NZ094P32S7C001	CY039974
		NIGSP-NZ-00094.7b	NZ094P11S7C005	CY039975
		NIGSP-NZ-00094.7c	NZ094P11S7C004	CY039976
		NIGSP-NZ-00094.7d	NZ094P11S7C007	CY039977
		NIGSP-NZ-00094.7e	NZ094P11S7C008	CY039978
		NIGSP-NZ-00094.7f	NZ094P32S7C002	CY039979
		NIGSP-NZ-00094.7g	NZ094P32S7C004	CY039980
		NIGSP-NZ-00094.7h	NZ094P32S7C006	CY039981
		NIGSP-NZ-00094.7i	NZ094P32S7C005	CY039982
		NIGSP-NZ-00094.7j	NZ094P32S7C007	CY039983
	8 (NS)	NIGSP-NZ-00094.8a	NZ094P32S8C001	CY039984
Influenza A virus New York/537/1998 (H3N2)	4 (HA)	NIGSP-NYS-00537.4a	WW537H3N2S4C34	CY039878
		NIGSP-NYS-00537.4b	WW537H3N2S4C8	CY039879
		NIGSP-NYS-00537.4c	WW537H3N2S4Cr1	CY039880
		NIGSP-NYS-00537.4d	WW537H3N2S4Cr5	CY039938
		NIGSP-NYS-00537.4e	WW537H3N2S4Cr7	CY039882
		NIGSP-NYS-00537.4f	WW537H3N2S4Cr8	CY039883
		NIGSP-NYS-00537.4g	WW537H3N2S4C36	CY039884
		NIGSP-NYS-00537.4h	WW537H3N2S4C6	CY039885

populations. As a large number of point mutations are thought to be required for an avian influenza virus such as A/H5N1 to evolve sustained transmission in human populations (5), one likely scenario for successful emergence is through the acqui-sition of genetic material from a viral subtype already adapted to humans, such as A/H1N1 or A/H3N2. This would require that viruses of both subtypes coinfect the same cells, thereby generating a mixed infection, and then exchange genomic seg-

a M segment
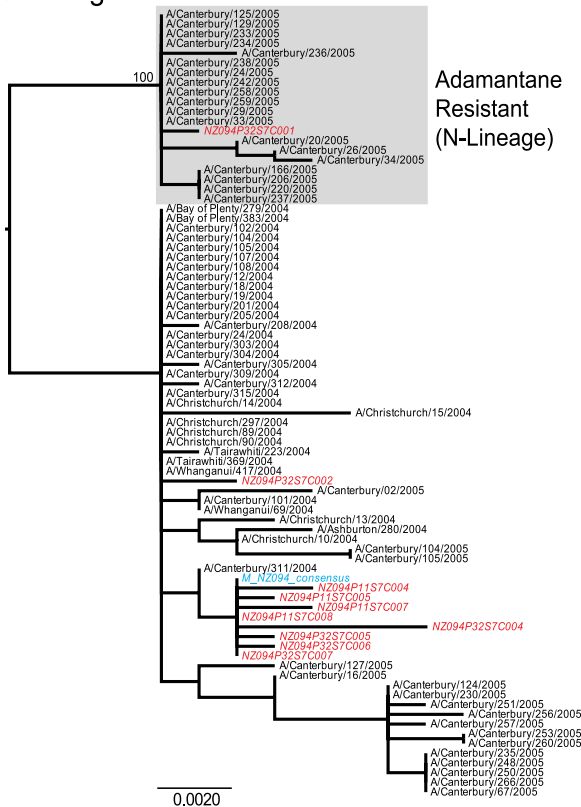
b PB1 Segment
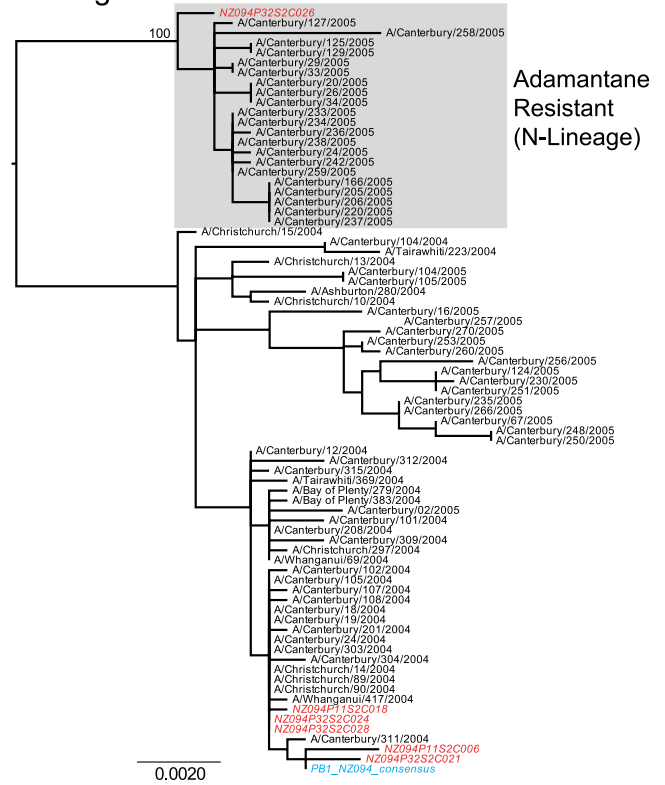
c PA Segment
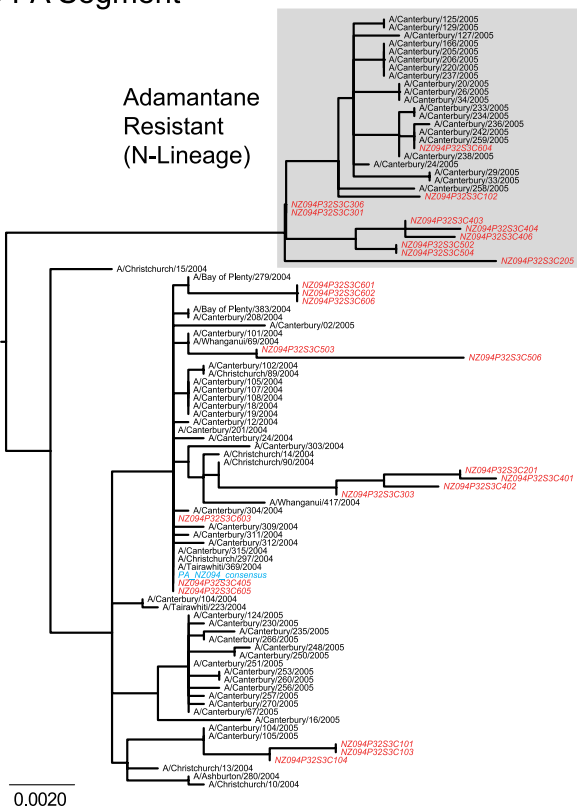
d NP Segment
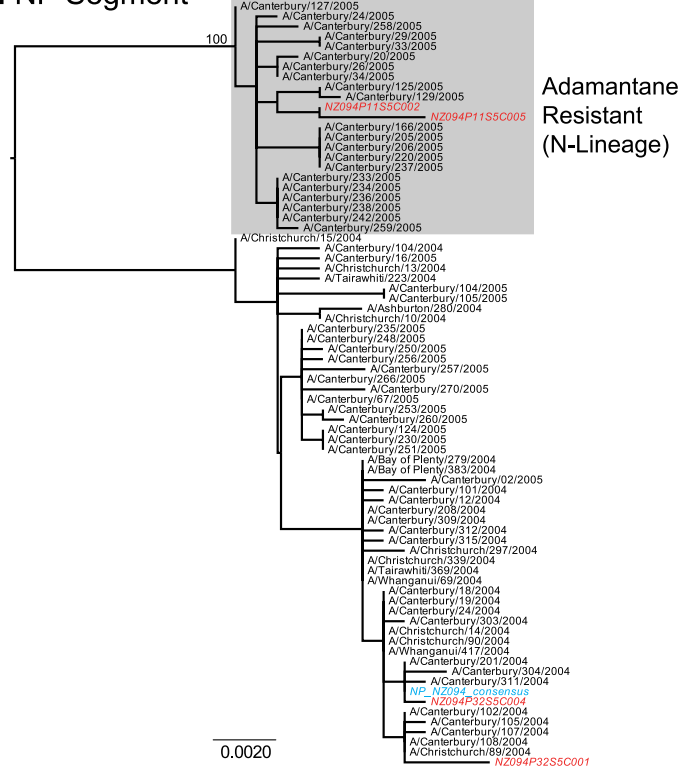
ments through reassortment, as was the case in 1957 and 1968. As a consequence, it is crucial to determine the frequency with which mixed infection naturally occurs in influenza A virus as well as its phenotypic consequences. To address these questions we undertook, for the first time, in-depth sequencing of multiple viral genome sequences sampled from individual influenza patients. These studies were performed with approval of the New York State (study numbers 04-103 and 02-054) and University of Pittsburgh (08-110400) institutional review boards.

MATERIALS AND METHODS

Virus isolation. The virus isolates used in this study were taken from repositories of human influenza virus samples collected as part of the surveillance program and diagnostic service provided by the Virus Reference and Surveillance Laboratory at the Wadsworth Center, Albany (New York State Department of Health) and the Canterbury Health Laboratories (New Zealand). Viruses were passaged minimally in primary rhesus monkey kidney (pRhMK) or Madin-Darby kidney (MDCK) cell cultures and the RNA was extracted from the clarified supernatant.

Cloning and sequencing. cDNA was produced from isolated RNA using Superscript III (Invitrogen) and a universal primer for all segments (Uni12, AGC AAAAGCAGG) (7). PCR was then performed with Expand high-fidelity polymerase (Roche) or AmpliTaq Gold (Applied Biosystems) using primers specific for each individual segment. Amplicons from genomic segments were cloned into a TOPO vector (Invitrogen), and individual clones were picked for sequencing using M13 and degenerate primers specific for regions along the segment sequence. Primer sequences are available on the J. Craig Venter Institute (JCVI) website (http://msc.jcvi.org/influenza).

Whole-genome sequencing of isolates was performed using the high-throughput sequencing pipeline at the JCVI. Briefly, an M13 sequence tag was added to the 5′ end of each degenerate primer used for sequencing. Primers are designed to produce approximately 500-nucleotide (nt) overlapping amplicons and 2× coverage of each genomic segment. Each primer pair overlaps with its neighboring primer pair by approximately 100 nt. Additionally, a second set of primers is designed to produce 500-nt amplicons offset by about 250 nt from the original primer pair, providing at least 4× sequence coverage of each segment. Primers were arranged in a 96-well plate format; all reverse transcription-PCRs (RT-PCRs) for each sample were performed in one plate. Genomic RNA was amplified directly by RT-PCR and sequenced. Ninety-six RT-PCRs were performed per RNA sample using a One-Step RT-PCR kit (Qiagen). Sequencing reactions were performed using Big Dye Terminator chemistry (Applied Biosystems) with 2 μl of template cDNA. Each amplicon was sequenced from each end using M13 primers, and sequencing reactions were analyzed on a 3730 ABI sequencer. Cloned segments were sequenced in a similar manner at the University of Pittsburgh Genomic and Proteomic Core Laboratories. After sequencing, the readouts were trimmed to remove amplicon primer sequence as well as low-quality sequence, and segments were assembled individually using the small-genome assembler Elvira (6). The clones are listed in Table 1.

Pyrosequencing. The primers used to amplify PCR products for pyrosequencing are the ones from the IGSP pipeline used for whole-genome sequencing. The primer sequences are available on the JCVI website (http://msc.jcvi.org /influenza). Three primer pairs were used to generate overlapping amplicons (459 to 522 nt in length) covering the HA1 region of the viral hemagglutinin (130F-589R, 391F-850R, and 453F-975R), one primer pair was used for the NA segment (380 nt; 180F-560R), and one primer pair was used for the M segment (450 nt; 478F-928R). Each amplicon was sequenced from either end. For the M segment, this resulted in two sets of pyrosequence reads that did not overlap. The primers were tagged with multiplex identifiers (Roche) (bar codes) so they could be sequenced with other samples and recognized individually. Products were

sequenced on the Roche GS-FLX pyrosequencing instrument at the University of Pittsburgh Genomic and Proteomic Core Laboratories. Samples were sequenced as part of a larger set on a four-well picotiter plate, and coverage was aimed to be 50 to 100 times per amplicon. The number of raw reads generated was 1,270 with an average length of 188 nt. The base-calling parameters were adjusted from their defaults to the most stringent values allowed in an effort to reduce the number of sequencing errors. Reads were trimmed (leading to an average length of 167 nt) to ensure that primers used in the PCR amplification step did not modify the nucleotide sequence. To further reduce the occurrence of sequencing errors in the data, a statistical error correction procedure based on that described by Eriksson and colleagues (4) was applied. A pairwise alignment of the reads to the consensus assembly obtained from whole-genome sequencing was performed using MUMmer (2), an algorithm which aligns and clusters matches.

Influenza A and influenza B virus real-time single-plex TaqMan RT-PCR assays. Viral RNA was extracted from 140 μl of primary swab or cell culture supernatant using Qiagen's QIAamp viral RNA kit (Valencia, CA) on a QIAcube (Qiagen). Five microliters of template RNA was added to 20 μl of master mix, and a quantitative one-step RT-PCR was performed using the qScript One-Step qRT-PCR kit of Quanta BioSciences (Gaithersburg, MD). Each assay was run on the Stratagene Mx3000P QPCR system (La Jolla, CA) for detection. Both assays are clinically approved for diagnostic use by New York State's Clinical Laboratory Evaluation Program. The influenza A virus assay targets the matrix protein (M), and the influenza B virus assay targets the nonstructural protein (NS). Both assays were developed in the Laboratory of Viral Diseases at the Wadsworth Center. Cycling conditions are as follows: 20 min at 48°C, 5 min at 95°C, and 45 cycles of 15 s at 95°C and 45 s at 55°C. Initial viral RNA copy numbers from extracted swabs and cultures were calculated from a standard curve compiled from real-time amplification of in-house-developed, amplicon-specific RNA transcripts previously quantitated by UV spectrophotometry. Many steps were taken to prevent cross-contamination of specimens in this study. First, to prevent contamination of the original specimens by the cultured isolates, the original swabs and corresponding cultures were extracted on two separate QIAcube instruments in two different laboratories. Second, negative extraction controls were included to detect cross-contamination events within each extraction. Third, to rule out contamination of the PCR reagents or cross-contamination on the PCR plate, water was included as a no-template control. Finally, RNA from the original specimens was added first on the PCR plate and capped before the addition of any cultured RNA material, positive control material, or amplicon-specific RNA transcripts. We are therefore confident that the results displayed below in Table 4 are due to influenza A and B virus mixed infections and not laboratory contamination.

Evolutionary analysis. Sequences were aligned manually and phylogenetic analysis was undertaken using the maximum likelihood (ML) method available in the PAUP* package and utilizing TBR branch swapping (version 4.0b10) (21). In each case the best fit model of nucleotide substitution (generally GTR+I+Γ₄) was determined using MODELTEST (13), and parameter values are available from the authors on request. A bootstrap resampling procedure was used to assess the support for individual nodes on the tree, utilizing 1,000 replicate neighbor-joining trees with evolutionary distances inferred under the ML substitution model.
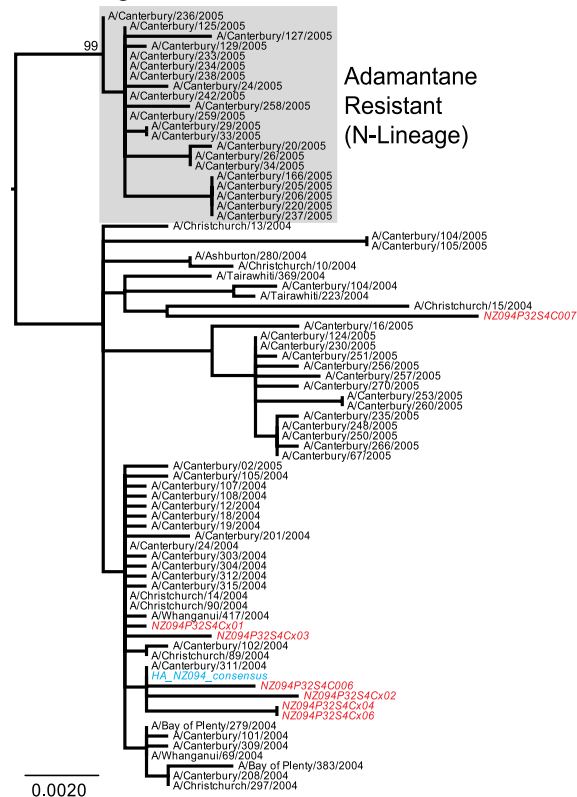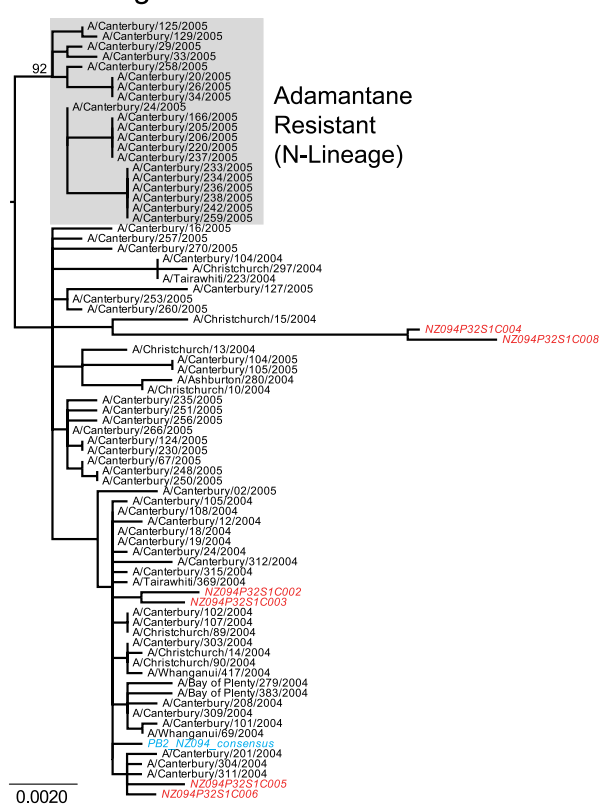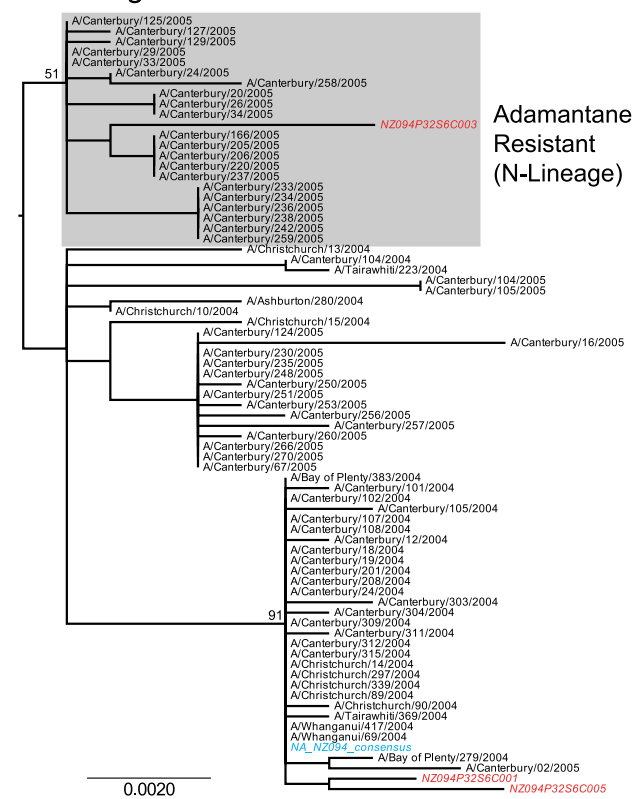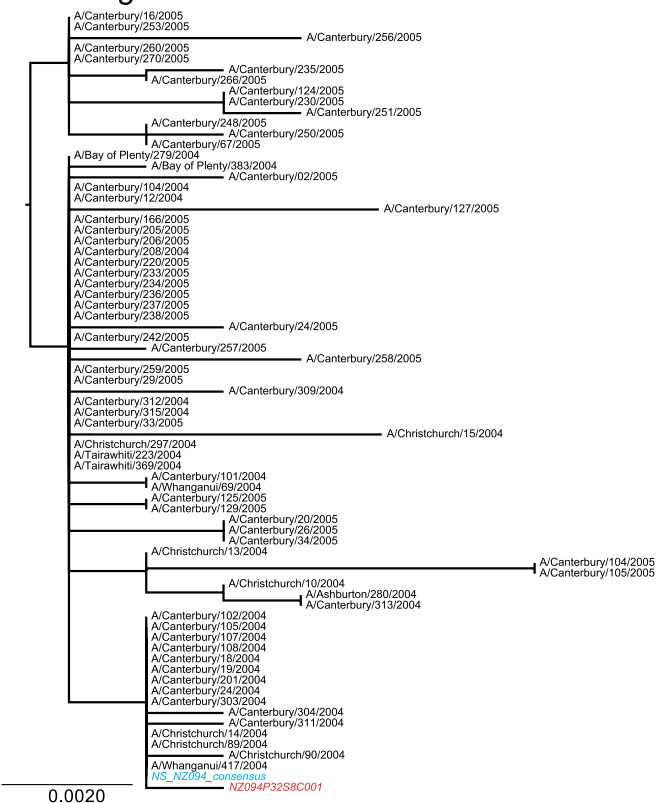
Nucleotide sequence accession numbers. Sequences for clones listed in Table 1 have been deposited in GenBank and assigned accession numbers CY039878 to CY039880, CY039882 to CY039892, and CY039938 to CY039984.

RESULTS

More than 1,250 samples of influenza A virus from New York State and New Zealand have been sequenced and published as part of the IGSP. In the IGSP standard sequencing pipeline all eight virus genomic segments of each isolate are

---

FIG. 1. Maximum-likelihood phylogenetic trees for M (a), PB1 (b), PA (c), and NP (d) segment clones from sample NZ094 compared to a representative sample of viruses circulating in the years 2004 and 2005. Clones from sample NZ094 are shown in red and italics, while consensus segment assemblies from whole-genome sequencing of the isolate are shown in light blue. Viruses associated with adamantane resistance are shaded. All branches are drawn to a scale of nucleotide substitutions per site, and the tree is midpoint rooted for purposes of clarity only. Bootstrap support values are shown for the node distinguishing the adamantane-resistant viruses. Only small (often nonoverlapping) subcloned regions were available for the PA segment, and so no bootstrap analysis could be undertaken in this case.

a HA segment

Adamantane
Resistant
(N-Lineage)

0.0020

b PB2 Segment

Adamantane
Resistant
(N-Lineage)

0.0020

c NA Segment

Adamantane
Resistant
(N-Lineage)
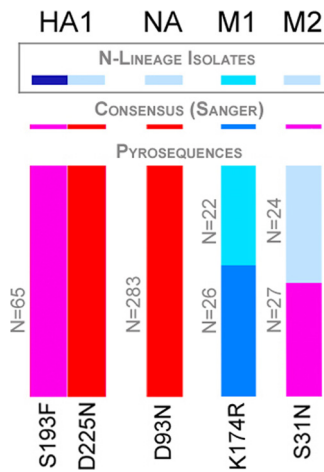
0.0020

d NS Segment

0.0020

FIG. 3. Color-coded representation of protein residues for translated pyrosequencing reads of the HA1, M, and NA segments of isolate NZ094. This representation shows the alignment of specific residues that differ from the N-lineage isolates A/Canterbury/20/2005 (GenBank accession numbers CY008340, CY008341, and CY008342), A/Canterbury/220/2005, (CY008380, CY008381, and CY008382), and A/Canterbury/237/2005 (CY008412, CY008413, and CY008414). These N-lineage isolates are highlighted by the gray box at the top of the figure. Protein residues from consensus assemblies from whole-genome sequencing of isolate NZ094 are represented above the pyrosequencing read data. The HA1 residues are numbered relative to the HA1 domain, not from the first methionine of the HA coding sequence. Color coding: light blue, N, asparagine; dark blue, F, phenylalanine; royal blue, R, arginine; turquoise, K, lysine; red, D, aspartic acid; pink, S, serine.

TABLE 2. Amino acid replacements characterizing the adamantane-resistant N-lineage that were found in cloned segments from patient NZ094

Protein	Amino acid replacement in N-lineage	NZ094	Clones (Sanger)[a]	Reads (pyrosequencing)[b]	% of reads
PB1	V113A	+	1/6		
PA	E101G	+	1/4		
	S208T	ND[c]	ND		
	K256Q	+	1/4		
	D382E	+	5/9		
	I421V	+	5/9		
	Y437H	+	6/10		
	I602V	+	1/3		
NP	A280V	+	2/4		
	I312V	+	2/4		
	S377G	+	2/4		
M1	K174R	+	1/9	22/48	46
	V219I	+	1/9	ND	
M2	S31N	+	1/9	24/51	47
HA1[d]	S193F	−	0/7	0/65	0
	D225N	−	0/7	0/65	0
NA	D93N	−	0/3	0/283	0

[a] Ratio of clones sequenced that possess amino acid replacement.
[b] Ratio of pyrosequence reads that possess amino acid replacement.
[c] ND, not determined.
[d] The residue positions for HA1 are given in HA1 numbering.

reconstructed from sequenced overlapping PCR products (or amplicons) and consensus assembly of the sequence reads. No cloning is involved. For 0.5% of the isolates sequenced, irresolvable base calls at multiple sites in genes indicated virus sequence heterogeneity, and full genomes could not be assembled. One of these isolates was collected in New Zealand in 2004, and its assembly led to a large number of unresolved ambiguities: A/Canterbury/200/2004 (abbreviated NZ094). We cloned each segment from this isolate to evaluate the extent of the genetic diversity present. An average of six clones per segment was sequenced by the Sanger method (Table 1).

Phylogenetic analysis of NZ094 clones revealed that at least two distinct lineages of A/H3N2 were present in this isolate: one closely related to viruses cocirculating in New Zealand during 2004 and a second lineage that clustered with A/H3N2 viruses that became dominant in the following (2005) influenza season in the southern hemisphere (Fig. 1). Notably, one of the nine clones of the M segment was closely related to 2005 viruses that carry the S31N mutation in the M2 ion channel protein and which confers resistance to adamantane drugs (18)

(Fig. 1A). The adamantanes (amantadine and rimantadine) normally block the M2 ion channel thus preventing the fusion of the virus and host-cell membranes. A single nonsynonymous point mutation (G to A at nucleotide 92 of the M2 open reading frame, leading to S31N) within the transmembrane region of this protein succeeds in producing a resistant virus.

Analysis of the other seven genome segments confirmed the presence of these two A/H3N2 lineages (Fig. 1 and 2). Specifically, a minority of the M, polymerase basic protein 1 (PB1), polymerase (PA), and nucleoprotein (NP) segment clones were more closely related to later 2005 viruses rather than to 2004 viruses (Fig. 1). In contrast, all clones for the other segments, HA, PB2, NA, and nonstructural protein (NS) either fell within the genetic diversity sampled in 2004 (HA and PB2) or possessed insufficient phylogenetic signal to clearly resolve evolutionary history (NA and NS) (Fig. 2). It is important to note that although two viral lineages must obviously be present in all segments, these are only clearly distinguishable in the PB1, PA, NP, and M segments.

This phenotypically important intrahost genetic diversity was

FIG. 2. Maximum-likelihood phylogenetic trees for the HA (a), PB2 (b), NA (c), and NS (d) segment clones from sample NZ094 compared to a representative sample of viruses circulating in the years 2004 and 2005. Clones from sample NZ094 are shown in red and italics, while consensus segment assemblies from whole-genome sequencing of the isolate are shown in light blue. Viruses associated with adamantane resistance are shaded in all cases. All branches are drawn to a scale of nucleotide substitutions per site, and the tree is midpoint rooted for purposes of clarity only. Bootstrap support values are shown for the node distinguishing the adamantane-resistant and -sensitive lineages. A grouping of adamantane-resistant viruses was only weakly supported in the NA gene (51% bootstrap support), and there was insufficient phylogenetic resolution in the NS gene to reveal any higher-order groupings.
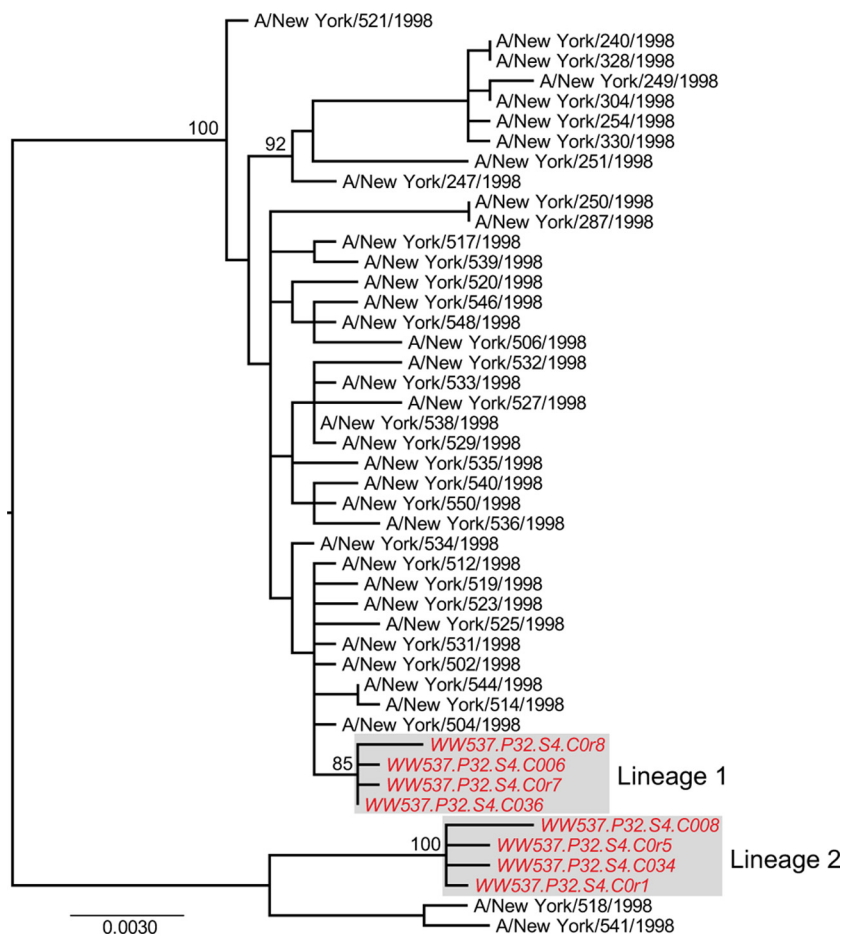
FIG. 4. Maximum-likelihood phylogenetic tree for HA segment clones from sample WW537 in comparison with those circulating during the same time period (1997–1998 northern hemisphere season) in New York State. Clones from sample WW537 are shown in red and italics. The two antigenically distinct lineages of HA are highlighted, and bootstrap support values are shown for key nodes. All horizontal branches are drawn to a scale of nucleotide substitutions per site, and the tree is midpoint rooted for purposes of clarity only.

confirmed through deep sequencing of specific regions of the NA, M, and HA1 domain of the HA derived from patient NZ094. Specific primers were used to amplify 250- to 400-nt regions of these segments, and amplicons were subjected to pyrosequencing on the GS-FLX (Roche/454). In both HA1 and NA, deep sequencing revealed no mutations characteristic of adamantane-resistant A/H3N2 (positions 193 and 225 for HA1 and position 93 for NA [18]) (Fig. 3). In marked contrast, a comparison with two positions in the M segment, one that corresponds to a residue in the M1 protein (K174R) and the other to the residue associated with drug resistance in M2 (S31N), shows the extent of the mixed variants (Fig. 3).

This complex pattern of genetic diversity indicates that patient NZ094 was carrying two phylogenetically distinct lineages of influenza A virus. That individual segments have such different evolutionary histories is strongly suggestive of mixed infection with already distinct lineages rather than generation de novo within this patient. That one lineage contains viruses that are adamantane resistant while the other does not further indicates that this patient was coinfected with the "parental" (i.e., prereassortant) strains of an adamantane-resistant lineage that later came to dominate global influenza virus diver-

sity following a major reassortment event in early 2005, denoted the N lineage (18). A detailed analysis of each clone for any of the 17 specific residues that were shown to be characteristic of the N lineage (18) further confirmed the presence of two separate lineages (Table 2).

The presence of cocirculating influenza A virus variants within one individual, with potentially important consequences for viral emergence, was also apparent in the case of a patient from New York State. Sample A/NewYork/537/1998 (abbreviated WW537) was collected during the 1997–1998 northern hemisphere influenza season. Phylogenetic analysis of eight fully sequenced HA clones (Table 1) revealed that these fell into two lineages separated by viruses sampled in other patients (Fig. 4), indicating that the genetic diversity within this patient was also generated through mixed infection rather than de novo. Notably, clones corresponding to these two lineages differed at multiple amino acid residues in the HA1 domain, suggesting that they are also antigenically distinct (Fig. 5). Focusing solely on the HA1 domain (residues 16 to 350, from the first methionine), 14 amino acid changes distinguish the two viral lineages, all falling at potential antibody binding sites. By comparing the HA1 domain of the clones to the A/H3N2
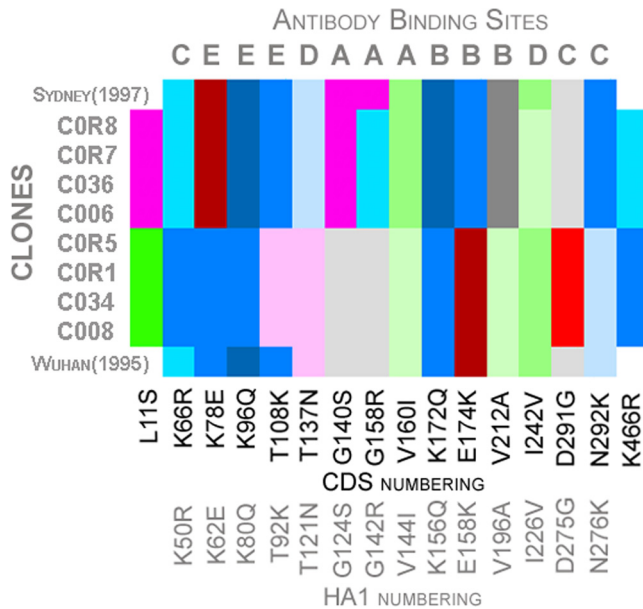
FIG. 5. Alignment of nonconserved HA1 residues of the HA across all eight clones from sample WW537. Numbering of the residues is given both from the first methionine of the coding sequence (CDS) and from the first residue of the HA1 domain (HA1 numbering). Gray letters at the top of each column correspond to residues of epitopic sites where antibodies have been shown to bind. L, leucine; S, serine; K, lysine; E, glutamic acid; Q, glutamine; T, threonine; N, asparagine; G, glycine; R, arginine; V, valine; A, alanine; I, isoleucine; D, aspartic acid.

strain used in the vaccine that season, A/Wuhan/359/1995, it is apparent that four of the clones (Cr08, C036, C006, and Cr07; lineage 1) consistently differ at 11 residues, again all representing antigenic residues. Indeed, these four clones are A/Sydney/05/97-like in appearance, such that they represent a drift variant of the A/Wuhan/359/1995 strain that spread rapidly in 1997 and to which the vaccine administered throughout North America during that year provided inadequate protection (1). A/Sydney/05/97 was chosen for the vaccine in the following year. This is the first reported case of a single patient being simultaneously infected with two cocirculating and antigenically distinct variants.

A final dramatic example of mixed infection was seen in another A/H3N2 patient from New York State—A/New York/347/1999 (WW347)—who was found to harbor both influenza A and influenza B viruses. Although this isolate was originally typed as influenza A virus (GenBank accession numbers AAZ74562 to AAZ74572), a complete influenza B virus genome was also recovered (GenBank accession numbers CY037367 to CY037374). No material remained from the original specimen, a primary swab collected in 1999. However, a portion of the harvest from the original pRhMK cell culture inoculated in 1999 was available for analysis. We therefore quantitated this sample alongside the subsequent, passaged cell culture isolate by real time-PCR. Diluted viral RNA from recently circulating 2006 influenza A/H1N1 and B virus strains, originally isolated in New York State, were used as positive amplification controls (Table 3). In New York State, this was the only case found of mixed influenza A and influenza B virus

infection out of the more than 500 isolates analyzed in the IGSP. After reviewing the more recent Wadsworth Center test records, however, two additional mixed influenza A/B virus coinfections were noted, and the relevant samples were retrieved for further analysis. Influenza A and B viral copy numbers were determined in both the original primary specimen and the first-passage isolate for both of these samples, using real-time RT-PCR as described above. The quantitation results confirm the presence of the two viruses in both of the primary swabs, as well as in the first passaged isolate in one of them (Table 4). No influenza B virus was detected in the other isolate, most likely due to the relatively low level of influenza B virus in the sample and consequent overgrowth of influenza A virus.

DISCUSSION

The vast majority of studies of genetic diversity in acute RNA viruses have been conducted on an epidemiological scale, in which a single, consensus sequence is obtained from each infected individual by direct population sequencing. This sequence by necessity then describes the most common variant in the intrahost viral population, thereby masking a myriad of mutant sequences. This is also true of the data generated under the Influenza Genome Sequencing Project. The IGSP was launched in 2004 to enhance the influenza virus genome knowledge base and provide the scientific community with complete genome sequence data for influenza A viruses, both in humans and other animal species. This project allowed an efficient viral genomics pipeline to be established to sequence various collections of human influenza virus from naso-pharyngeal specimens, avian influenza virus from wild bird rectal swabs, and swine influenza virus from nasal lavage samples. More than 3,000 full influenza virus genomes have been published since the project began, the majority of which have been used to explore the epidemiological-scale diversity of isolates assigned to the same subtype (8, 12). Indeed, a number of important and broad-scale conclusions have been drawn from the consensus sequence data generated by the IGSP, including the following: (i) multiple lineages of A/H3N2 and A/H1N1 viruses cocirculate during a single season, indicative of multiple entries into specific geographic regions (6); (ii) these lineages undergo frequent reassortment which may in turn have a major impact on antigenic evolution (8); (iii) these lineages

TABLE 3. Influenza A and B viral RNA copy numbers from isolate WW347, as determined using quantitative real-time RT-PCR

Sample	Viral RNA copy no./reaction mixture[a]	
	Influenza A virus	Influenza B virus
WW347 first pass	6.82E+07	9.24E+05
WW347 culture	3.99E+08	7.07E+07
No-template control	0	0
Amplification control[b]	2.20E+01	1.00E+02

[a] Values represent RNA extracted from a first-pass culture of a nasopharyngeal specimen and subsequent cell culture.
[b] The amplification control for influenza A virus was diluted RNA, extracted from an amplified New York A/H1N1 (2006) strain. The amplification control for influenza B virus was diluted RNA, extracted from an amplified New York B (2006) strain. Positive controls were usually prepared at a relatively low concentration.

TABLE 4. Influenza A and B viral RNA copy numbers from two samples, as determined using quantitative real-time RT-PCR

Sample ID no. (yr of isolation)	Sample type	Influenza A virus subtype	Viral RNA copy no./reaction mixture[a]	
			Influenza A virus	Influenza B virus
1 (2005)	Throat swab	H3N2	1.21E+06	6.80E+02
1 (2005)	RhMK1 culture	H3N2	2.21E+07	6.71E+07
2 (2009)	Pharyngeal swab	H1N1	4.21E+05	1.00E+01
2 (2009)	RhMK1 culture	H1N1	1.56E+07	0
Negative extraction control	PBS		0	0
No-template control	Water		0	0
Amplification control[b]	Viral RNA		2.00E+03	1.00E+02

[a] Values represent RNA extracted from the original specimen and first-pass cell culture.
[b] The amplification control for influenza A virus was diluted RNA extracted from an amplified New York A/H1N1 (2006) strain. The amplification control for influenza B virus was diluted RNA extracted from an amplified New York B (2006) strain.

experience complex global dynamics, suggestive of a source-sink ecological model (12, 14, 15). More recently, a large-scale survey of genomic data in wild birds revealed similarly complex dynamics in avian influenza virus, with extremely high rates of reassortment and some evidence of interhemisphere viral traffic (3, 16). Of particular significance was the observation that approximately 25% of all the isolates studied may have experienced mixed infection, highlighting how frequently this process occurs in the avian reservoir population.

Although consensus sequencing is valuable for many aspects of molecular epidemiology, it by necessity cannot shed light on evolutionary processes that take place within individually infected hosts. Our study has revealed a remarkably high level of mixed infection in human influenza virus, including diverse lineages of the same subtype that differ in both their propensity for drug resistance and antigenicity, and lineages of different virus types (A and B). A screening of the first 2,000 influenza virus samples published on GenBank for the IGSP shows that approximately 3% have some evidence of large-scale sequence polymorphism suggestive of mixed infection (unpublished observation). As published consensus sequences are necessarily skewed toward the dominant strain within isolates, this number is almost certainly a major underrepresentation of the true level of mixed infection. For example, we also found the clear signature of an H3N2/H1N1 mixed infection in an isolate (A/Canterbury/247/2005) sampled in New Zealand in 2005, although the primary specimen was no longer available for analysis and we therefore could not confirm that this coinfection was present in the host.

In sum, we propose that mixed infection of diverse influenza viruses, a necessary precursor to reassortment, is a common occurrence during seasonal influenza in humans and will in turn accelerate the rate of adaptive evolution in this virus. In addition, intrahost populations of influenza virus will harbor genetic diversity generated by de novo mutation, which we have not assessed in the current study. As a consequence, we urge that intrahost sequencing be more routinely employed to assess the degree of genotypic and phenotypic diversity in populations of acute RNA viruses. With the advent of high-throughput next-generation sequencing platforms, viral variants are being much more explicitly revealed within specimens, and this type of data can be made available on a routine basis.

REFERENCES

1. **Centers for Disease Control and Prevention.** 1998. Update: influenza activity—United States and worldwide, 1997–98 season, and composition of the 1998–99 influenza vaccine. MMWR Morb. Mortal. Wkly. Rep. **47:**280–284.
2. **Delcher, A. L., S. L. Salzberg, and A. M. Phillippy.** 2003. Using MUMmer to identify similar regions in large sequence sets. Curr. Protoc. Bioinformatics **10:**10.3.
3. **Dugan, V. G., R. Chen, D. J. Spiro, N. Sengamalay, J. Zaborsky, E. Ghedin, J. Nolting, D. E. Swayne, J. A. Runstadler, G. M. Happ, D. A. Senne, R. Wang, R. D. Slemons, E. C. Holmes, and J. K. Taubenberger.** 2008. The evolutionary genetics and emergence of avian influenza viruses in wild birds. PLoS Pathog. **4:**e1000076.
4. **Eriksson, N., L. Pachter, Y. Mitsuya, S. Y. Rhee, C. Wang, B. Gharizadeh, M. Ronaghi, R. W. Shafer, and N. Beerenwinkel.** 2008. Viral population estimation using pyrosequencing. PLoS Comput. Biol. **4:**e1000074.
5. **Finkelstein, D. B., S. Mukatira, P. K. Mehta, J. C. Obenauer, X. Su, R. G. Webster, and C. W. Naeve.** 2007. Persistent host markers in pandemic and H5N1 influenza viruses. J. Virol. **81:**10292–10299.
6. **Ghedin, E., N. A. Sengamalay, M. Shumway, J. Zaborsky, T. Feldblyum, V. Subbu, D. J. Spiro, J. Sitz, H. Koo, P. Bolotov, D. Dernovoy, T. Tatusova, Y. Bao, K. St George, J. Taylor, D. J. Lipman, C. M. Fraser, J. K. Taubenberger, and S. L. Salzberg.** 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature **437:**1162–1166.
7. **Hoffmann, E., J. Stech, Y. Guan, R. G. Webster, and D. R. Perez.** 2001. Universal primer set for the full-length amplification of all influenza A viruses. Arch. Virol. **146:**2275–2289.
8. **Holmes, E. C., E. Ghedin, N. Miller, J. Taylor, Y. Bao, K. St George, B. T. Grenfell, S. L. Salzberg, C. M. Fraser, D. J. Lipman, and J. K. Taubenberger.** 2005. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLoS Biol. **3:**e300.
9. **Horimoto, T., and Y. Kawaoka.** 2005. Influenza: lessons from past pandemics, warnings from current incidents. Nat. Rev. Microbiol. **3:**591–600.
10. **Lindstrom, S. E., N. J. Cox, and A. Klimov.** 2004. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. Virology **328:**101–119.
11. **Nayak, D. P., E. K. Hui, and S. Barman.** 2004. Assembly and budding of influenza virus. Virus Res. **106:**147–165.
12. **Nelson, M. I., L. Simonsen, C. Viboud, M. A. Miller, J. Taylor, K. S. George, S. B. Griesemer, E. Ghedin, N. A. Sengamalay, D. J. Spiro, I. Volkov, B. T. Grenfell, D. J. Lipman, J. K. Taubenberger, and E. C. Holmes.** 2006. Stochastic processes are key determinants of short-term evolution in influenza A virus. PLoS Pathog. **2:**e125.
13. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
14. **Rambaut, A., O. G. Pybus, M. I. Nelson, C. Viboud, J. K. Taubenberger, and E. C. Holmes.** 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature **453:**615–619.
15. **Russell, C. A., T. C. Jones, I. G. Barr, N. J. Cox, R. J. Garten, V. Gregory, I. D. Gust, A. W. Hampson, A. J. Hay, A. C. Hurt, J. C. de Jong, A. Kelso, A. I. Klimov, T. Kageyama, N. Komadina, A. S. Lapedes, Y. P. Lin, A. Mosterin, M. Obuchi, T. Odagiri, A. D. Osterhaus, G. F. Rimmelzwaan,

M. W. Shaw, E. Skepner, K. Stohr, M. Tashiro, R. A. Fouchier, and D. J. Smith. 2008. The global circulation of seasonal influenza A (H3N2) viruses. Science **320:**340–346.

16. Salzberg, S. L., C. Kingsford, G. Cattoli, D. J. Spiro, D. A. Janies, M. M. Aly, I. H. Brown, E. Couacy-Hymann, G. M. De Mia, H. Dung do, A. Guercio, T. Joannis, A. S. Maken Ali, A. Osmani, I. Padalino, M. D. Saad, V. Savic, N. A. Sengamalay, S. Yingst, J. Zaborsky, O. Zorman-Rojs, E. Ghedin, and I. Capua. 2007. Genome analysis linking recent European and African influenza (H5N1) viruses. Emerg. Infect. Dis. **13:**713–718.

17. Scholtissek, C., W. Rohde, V. Von Hoyningen, and R. Rott. 1978. On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology **87:**13–20.

18. Simonsen, L., C. Viboud, B. T. Grenfell, J. Dushoff, L. Jennings, M. Smit, C. Macken, M. Hata, J. Gog, M. A. Miller, and E. C. Holmes. 2007. The genesis

and spread of reassortant human influenza A/H3N2 viruses conferring adamantane resistance. Mol. Biol. Evol. **24:**1811–1820.

19. Wiley, D. C., and J. J. Skehel. 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu. Rev. Biochem. **56:**365–394.

20. Wiley, D. C., I. A. Wilson, and J. J. Skehel. 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature **289:**373–378.

21. Wilgenbusch, J. C., and D. Swofford. 2003. Inferring evolutionary trees with PAUP*. Curr. Protoc. Bioinformatics **6:**6.4.

22. Wilson, I. A., J. J. Skehel, and D. C. Wiley. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature **289:**366–373.